

Efficiency analysis of brokers in the electronic marketplace

Virgílio A.F. Almeida^{a,1,2}, Wagner Meira Jr.^{a,*,1,3}, Victor F. Ribeiro^{b,4}, Nivio Ziviani^{a,b,1,5}

^a Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

^b Miner Technology Group, Belo Horizonte, MG, Brazil

Abstract

In this paper we analyze the behavior of e-commerce users based on actual logs from two large non-English e-brokers. We start by presenting a quantitative study of the behavior of e-brokers and discuss the influence of regional and cultural issues on them. We then discuss a model that quantifies the efficiency of the results provided by brokers in the electronic marketplace. This model is a function of factors such as server response time and regional factors. Our findings clearly indicate that e-commerce is strongly tied to local language, national customs and regulations, currency conversion and logistics, and Internet infrastructure. We found that the behavior of customers of online bookstores is strongly affected by brand and regional factors. Music CD shoppers show a different behavior that might stem from the fact that music is universal and not so language dependent. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Electronic commerce; World Wide Web; Internet; E-brokers

1. Introduction

The Internet and the World Wide Web provide a global virtual marketplace, without location and time constraints. The electronic market provided by the almost universal system of communication of the Web is adequate for information-based products (e.g., news, software, financial services, ticketing services) and also for order retailing of some non-digital products such as books, CDs, flowers, travel, groceries, PCs, among others. Usually, e-commerce

companies make available on the Internet electronic catalogs, that support lists of products and/or services, price information, and commercial transactions. As a consequence, the amount of available information and the number of potential customers in the Web is growing very rapidly [11].

Though useful information may exist somewhere, it is not always easy to find what a user is looking for on the Web. Since the Web is large and growing exponentially, it is impractical to exhaustively browse the Web looking for products and services. Therefore, one of the biggest challenges faced by electronic customers is the information overload, that hampers the growth of the online buying process. Although there are several different models for representing e-customer behavior, there exist some basic steps that are shared by most models [2], such as: need identification, product search, merchant search, negotiation, purchase and delivery,

* Corresponding author.

¹ Partially supported by Project SIAM-DCC-UPMG, grant MCT-PINEX-PRONEX number 76.97.1016.00.

E-mail: {virgilio,meira,nivio}@dcc.ufmg.br

² Partially supported by CNPQ grant 300437/87-0.

³ Partially supported by CNPQ grant 380134/97-7.

⁴ E-mail: victor@miner.com.br

⁵ Partially supported by CNPQ grant 520916/94-8.

product service and product evaluation. In order to boost e-commerce activities, tools and services are needed to help customers in each of these basic steps.

As a result, e-brokers have been developed to help users to find information, products and merchants. A broker is a party which mediates between buyers and sellers in a marketplace [16]. E-brokers can search for products, retrieving information to help a customer to determine what to buy. E-brokers can also look for merchant-specific information (e.g., price) to help a customer to decide whom to buy from. Basically, e-brokers can be viewed as search engines that specialize in specific topics. For example, a bargain broker searches the Web for price and characteristics of the products, summarizes the results and presents it to the user. In another example, a broker could search e-catalogs of many suppliers, which are registered with the broker, and try to match product specification and negotiation requirements.

The Miner Family of Web Agents [13] is a set of tools whose main objective is to help people in finding information on the Web. The main idea is to bring multiple search and information sources together in one place. The searching is performed by agents working in parallel, just like metasearchers [15,4] that use several search engines simultaneously, collecting answers and unifying them. The information may be the price of a book, a new musical release, a freeware or a shareware software, daily news, or any document available on the Web.

A *portal* is a site that brings together a variety of content and services in one area and attracts a large number of visitors. The idea is to become the single best starting place for as many users as possible. In Brazil, the largest Web site is UOL [17], which is shaping itself as a portal. UOL is a Brazilian site that brings together a variety of content and services in different areas. UOL acts both as a content and service provider offering more than 53 Brazilian magazines, 21 international magazines, 59 Brazilian newspapers, and 31 international newspapers. UOL also offer several services, including hundreds of chat rooms that topped 12,000 people online simultaneously, more than 400 product sites, and RadarUOL, a search engine powered by Inktomi [6]. UOL topped more than 12 million page views

in one day⁶, while being one of the largest non-English and the largest Portuguese content provider in the world. The Miner Family has a partnership deal with UOL and is one of the services offered at the UOL site. The Miner Family has topped more than 100,000 page views in one day. This rich environment provided us the data used in this paper.

The goal of the paper is threefold. First, we give an overview of the Miner Family architecture, implementation and workload characteristics and point out the differences to existing similar services (e.g., Express, Jango, and Junglee). Second, we present a quantitative study of the behavior of two large non-English e-brokers. Considering that the e-brokers are browsed by a large number of users who mainly speak Portuguese and live in Brazil, we discuss the influence of regional and cultural issues on the e-commerce activities. Third, we present a case study that analyzes the efficiency of the results provided by e-brokers in the e-marketplace.

The paper is organized as follows. Section 2 presents the architecture of the Miner Family and discusses its design rationale, components, and overall workload. Section 3 characterizes the workload of two brokerage services (i.e., BookMiner and CD-Miner) of the Miner Family. Section 4 presents a click-through model based on data collected from the operation of the e-broker BookMiner. We present figures that indicate the level of activity of the e-broker and show a model of the customer behavior. To conclude, Section 5 points out some evidences that show the influence of regional and cultural issues (language in particular), brand and regional factors, on the quantitative results presented in the paper.

2. Architecture and workload of the Miner Family

The Miner Family of Web Agents [13] is both a searching utility and an electronic catalog, that also provides brokerage services. The Miner Family was developed mainly for Portuguese language-

⁶For sake of comparison, Pathfinder and the CNN group had about 3 million and 14 million page views per day in the same period.

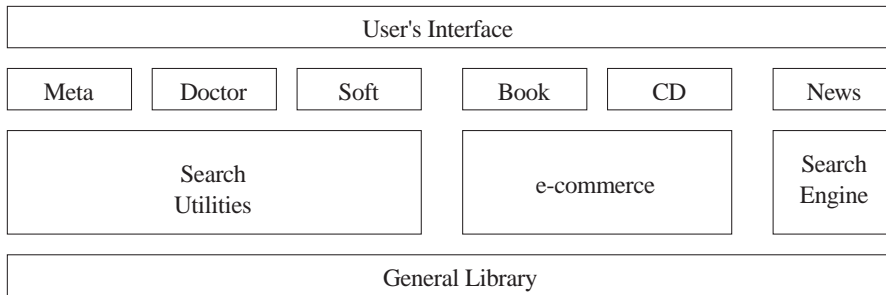


Fig. 1. Structure of the Miner Family code.

based services. The search utility services provided by the Miner Family at the time the paper was written include: (1) MetaMiner, metasearch engine that uses Brazilian and international search engines, (2) DoctorMiner, that searches for information on several sites containing medical and dental articles, (3) SoftMiner, that searches for software in freeware and shareware sites, the just released (4) JavaMiner, that searches for technical information about Java language, and (5) PeopleMiner, that searches for people on the Internet. The search engine service includes (6) NewsMiner, that collects news from Brazilian newspapers, leaving them daily available for the Internet community. Brokerage services include: (7) BookMiner, that searches for books in registered Brazilian and international bookstores to match user's specification and (8) CDMiner, that searches for musical titles in Brazilian and international music stores to find the user's preferences. Table 1 presents description of each member concerning its target (e.g., search engines, stores, etc.) and the number of registered sites for each member.

Table 1
Members of the Miner Family

Member	Target	#Sites
MetaMiner	search engines	13
DoctorMiner	medical and odontological information	17
NewsMiner	newspapers	13
BookMiner	bookstores	16
CDMiner	music stores	13
SoftMiner	software	10
PeopleMiner	people	13
JavaMiner	Java language information	7

The Miner Family was coded in Java and comprises about 23,000 lines of code that run on a Netscape Enterprise Server, and the host platform is a SUN Ultra running Solaris 2.6. The code was implemented emphasizing greater reusability and easier maintenance and is structured into four levels: (1) general library, (2) middleware (e-commerce, search utilities, and search engines), (3) agents, and (4) user's interface. Fig. 1 depicts each of these levels, which are explained in detail in the next paragraphs.

The general library contains several functionalities that are used by the upper levels, such as handlers (HTTP, cookies, tickets), query caching (for breaking results among pages), data fusion and interface widgets. It corresponds to 25% of the Miner code. The functions and primitives for each of the types of services offered by the Family are implemented in the middleware level, and each of the three services comprises about 2000 lines of code. The e-commerce code contains classes that abstract goods' characteristics and interface with the stores that sell them. Similarly, the search utilities code contains functions that handle searches in each of the types of sites (software, people, and general) and the respective object classes. The search engines code implements procedures that follow the ethic of 'bots' [3,9], an information manager, connection handler, and bots' navigation control. The agents responsible for querying the various sites comprise 3000 lines of code total. Among other tasks, these classes store details about site handling, data filtering, and the structure of HTML data. Finally, the interface code (7000 lines) implements all the HTML forms for the queries and the formatting of their results.

By using this structure, the implementation of new Family members becomes trivial. A new search

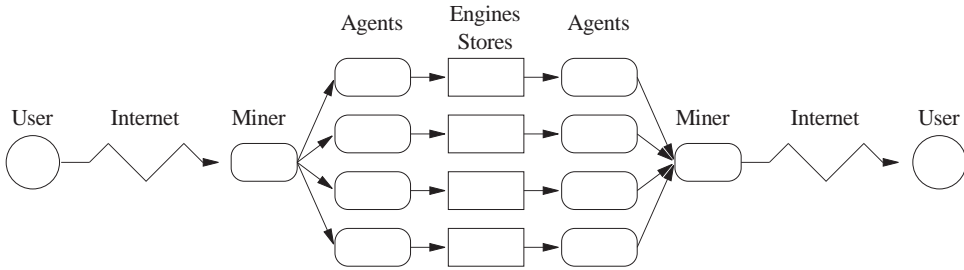


Fig. 2. Miner Family functionality.

utility, querying ten different sites would require only about 500 new lines of code. As an example, the implementation of the newest member, JavaMiner, cost 16 man-hours and was made available in less than a week after conception.

All members of the Miner Family work similarly and the main steps to answer a query are depicted in Fig. 2. Each query task can be divided into five main steps, as follows: (1) a user submits a query; (2) the Miner server gets the query and dispatches its agents; (3) each agent queries its target engine, store, or site; (4) each agent receives and parses the query results; and (5) the server unifies, formats, and sends the results back to the user.

2.1. Workload characterization of the Miner Family

This section presents a workload characterization of the Miner Family. We start out the analysis by partitioning the overall workload according to the services provided by the Miner Family. Table 2 shows the data extracted from logs of a four-week period of usage of the Miner services. The daily average number of requests was 22,086. We divided

the data into three categories: (1) request frequency, (2) request characteristics, and (3) hourly distribution. Request frequency represents the percentage of requests addressed to each service. We note that MetaMiner is the most popular service, receiving almost 90% of the total requests. Three other metrics were defined to further characterize the request workload: (1) words per query, (2) match ratio, and (3) answers per query. Words per query quantifies the complexity of the request, which is around 2 words on the average. For instance, 95% of the requests to CDMiner have less than four words.

The match ratio represents the number of requests that returned at least one URL. In this case, we can observe that a high match ratio can result from two different scenarios. The first one is related to services that have broad coverage (i.e., the MetaMiner) and provide answers for most of the queries (although we cannot quantify how meaningful the answers are). The second scenario involves services that are so specialized that the queries are very constrained (i.e., SoftMiner and DoctorMiner). Similar conclusions arise when we look at the average number of answers per query.

Regarding hourly distribution, we consider three characteristics of the workload: peak period, peak hour, and peak/average ratio. Peak period represents the hours during which the number of requests is higher than the daily average. As we can see in Table 2, this information uncovers an interesting characteristic of Miner users, who usually query information during work time, probably using a non-modem connection. The peak hour is the time slot when the maximum number of requests was observed. In all cases but two, we noticed the peak hour is during lunch time in Brazil. One of the exceptions occurs for the NewsMiner service, whose

Table 2
Overall workload statistics

Miner	Meta	Book	CD	Soft	News	Doctor
Queries (%)	89.15	2.60	2.65	2.34	1.89	1.37
Words/Query	1.98	2.05	1.87	1.55	1.66	1.69
Match Ratio (%)	93.64	75.65	79.53	88.00	55.60	95.81
Answers/Query	53.97	42.40	41.06	63.74	11.05	47.78
Peak Period	7 am	7 am	11 am	8 am	5 am	8 am
	9 pm	9 pm	7 pm	11 pm	5 pm	10 pm
Peak Hour	1 pm	1 pm	2 pm	1 pm	7 am	8 pm
Peak Ratio	2.29	7.52	6.41	7.50	13.12	9.37

peak is around 7:00 am, when users log to get the daily and breaking news. DoctorMiner peak hour is around 8:00 pm, when health professionals are usually able to look for medical information. Finally, peak ratio measures the request rate at the peak hour over the average rate [12]. Specific services such as BookMiner and NewsMiner are more bursty than generic search services like MetaMiner. Their peaks are 7 and 13 times higher than their average, respectively, while the MetaMiner peak ratio is only 2.29.

2.2. Related work

There are related works in this area. Excite has a shopping guide to find products and prices on the Web, which is called Product Finder and is powered by Jango [7]. Junglee [8] has developed a technology which aggregates information and prices for merchandise sold on the Web, enabling consumers to compare and shop for online products. Their technology is now being used by Yahoo [18,19]. More recently, Infoseek announced Express [5], which uses many search engines to multiple search for products.

Table 3 presents the main characteristics of the three technologies mentioned above and the Miner Family. The first row shows the number of bookstores used by Yahoo.Junglee, Infoseek.Express and BookMiner. In the case of Yahoo.Junglee the number was estimated from the queries submitted as they do not list the actual bookstores. In the case of Infoseek.Express they do not search all five bookstores or music stores in parallel, but each one at a time, and so we could not include them in our experiment whose results are shown in Table 4. From the 16 bookstores listed in BookMiner 8 are Brazilian.

Table 3
Characteristics of the search tools

Characteristics	Technologies			
	Junglee Yahoo	Jango Excite	Express Infoseek	Miner
Bookstores	6 ^a	–	5	16 (8 Brazilian)
Music stores	4 ^a	–	5	13 (5 Brazilian)
Software	–	10	5	10
Metasearch engines	–	–	7	13 (5 Brazilian)
Parallel search	yes	yes	no	yes
Where to search option	no	no	yes	yes

^a Estimated.

Table 4

Different types of queries submitted to Yahoo.Junglee and the equivalent Miner tools (BookMiner and CDMiner)

Queries	Answers		
	Junglee	Miner	Common
Sphere (by title)	75	261	65
Jurassic Park (by title)	71	106	58
Tom Wolfe (by author)	77	46	40
Fernando Pessoa (by author)	30	160	27
Jorge Amado (by author)	39	225	35
Ella Fitzgerald (by artist)	42	161	20
Joao Gilberto (by artist)	28	76	11
Subway (by title)	0	1	0

The second row presents the number of music stores provided by Yahoo.Junglee and CDMiner, and the value for Yahoo.Junglee is again estimated from the queries because they do not list them. From the 13 music stores listed in CDMiner 5 of them are Brazilian ones. The third row presents the number of engines to search for software (freeware and software). Again, Infoseek.Express searches all 5 software sites one at a time. The fourth row presents the number of search engines and directories used by Infoseek.Express and MetaMiner. From the 13 engines used by MetaMiner 5 are Brazilian. The fifth row shows that only Infoseek.Express searching tools do not perform requests in parallel. Finally, the last row shows the tools that allow users to choose the sites that are to be queried.

Table 4 presents seven different queries submitted to Yahoo.Junglee, BookMiner and CDMiner. The first five queries search for books, the first two being titles published in US. The following three

are authors of books: one American (i.e., the writer Tom Wolfe), one Portuguese (i.e., the poet Fernando Pessoa), and one Brazilian (i.e., the writer Jorge Amado). The following two queries search for CDs from one American artist (i.e., the jazz singer Ella Fitzgerald) and one Brazilian artist (i.e., the bossa nova singer João Gilberto). The last query searches for the sound track of the 1985 movie *Subway*, which was found only in one Brazilian music store at that time. The aforementioned table shows the query results. The first two columns present the answers returned by Junglee and Miner, respectively. The last column (Common) presents the number of answers that appeared in the results returned by both tools. The large number of documents returned by the Miner Family comes from the larger number of registered sites. For queries involving Brazilian and Portuguese names the differences are even larger because of the language influence.

3. Workload of the two e-brokers

This section analyzes the workload of two brokerage services, namely BookMiner and CDMiner. The goal is to study the actual workload generated by customers searching for books and CDs on global and Brazilian electronic stores. The characterization is based on data collected from two logs, corresponding to a four-week period. The first log shows overall results of the broker activities, while the second one provides per-store information. IP addresses were masked in order to protect users' privacy. We merged the two logs based on time, date, and masked IP address. As a result, the merged logs provide the following information: date and time of the request, query keyword(s), type of query (title or author), request response time, overall number of titles or CDs returned to the user, response time for each store, and number of titles or CDs returned by each store.

The broker workload is described by a graph called a *Customer Preference Graph (CPG)*. This graph has one node for each service and registered stores of the broker. The transitions between the nodes represent the percentage of customers that followed a specific path, i.e., service, national domain, and store. Fig. 3 shows the CPG for the BookMiner brokerage service. For each registered

bookstore, we measured the click-through frequency, given a BookMiner response. The click-through determines which bookstore was chosen by the user. The percentage associated with each path of the BookMiner graph represents the click-through frequency. From the CPG of Fig. 3, we note that 76% of the customers prefer Brazilian bookstores. Among the global bookstores, Amazon.com was chosen by most of the users (50% of the users), followed by Barnes & Noble and BookStacks. Siciliano, a Brazilian bookstore, is responsible for one fourth of the click-throughs among the Brazilian bookstores, followed by Cultura, Booknet and Loyola.

On the other hand, the CPG of the CDMiner (Fig. 4) shows a different customer profile. The percentage of users that visit global and Brazilian stores are about the same. We conjecture that this behavior can be explained by the following observations: (1) international music has wider acceptance than international literature in Brazil, and (2) no Brazilian music stores allow consumers to listen to tracks from CDs before buying. According to [14], customers of stores that sell music CDs want more music samples, ease of use and low prices. In addition to the fact that Brazilian music stores do not offer samples, the tax to import CDs may explain why customers visit international stores, but do not buy the products. On the contrary, books do not pay import tax. Among the global bookstores, Blockbuster was chosen by 26.77% of the users, followed close by CDNow and Amazon.com. It is remarkable that CDNow, as an electronic music store, is more famous than BlockBuster. We conjecture that two factors led shoppers to visit BlockBuster more frequently: (1) it does not return the prices of the CDs, which somehow forces customers to visit its site, and (2) the average response time of CDNow is four times larger than Blockbuster, as we discuss later. CDSstudio, a Brazilian music store, is responsible for almost one third of the click-throughs among the Brazilian music stores, followed by Ferrs. Again we can observe how specialization affects user preferences, VanDamme, which sells only *new age* CDs, got only 1.21% of the click-throughs.

E-commerce service levels are usually assessed through response time and availability. For the purpose of our analysis, a server is considered available when it answers the book request within the user-defined timeout (i.e., 60 seconds by default). Elapsed

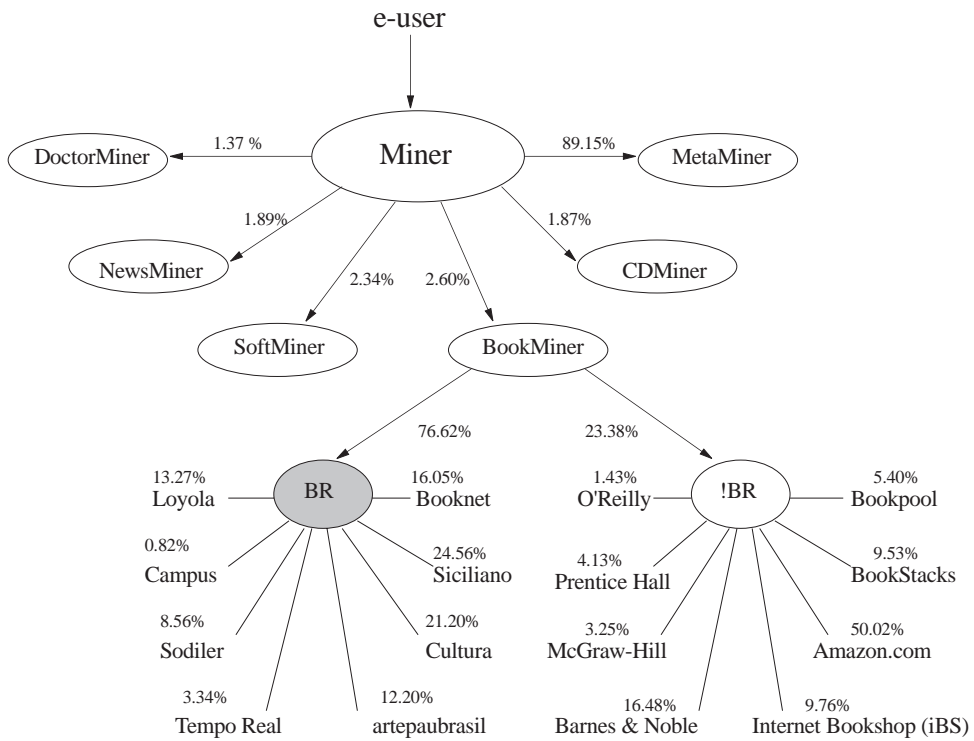


Fig. 3. BookMiner customer behavior graph.

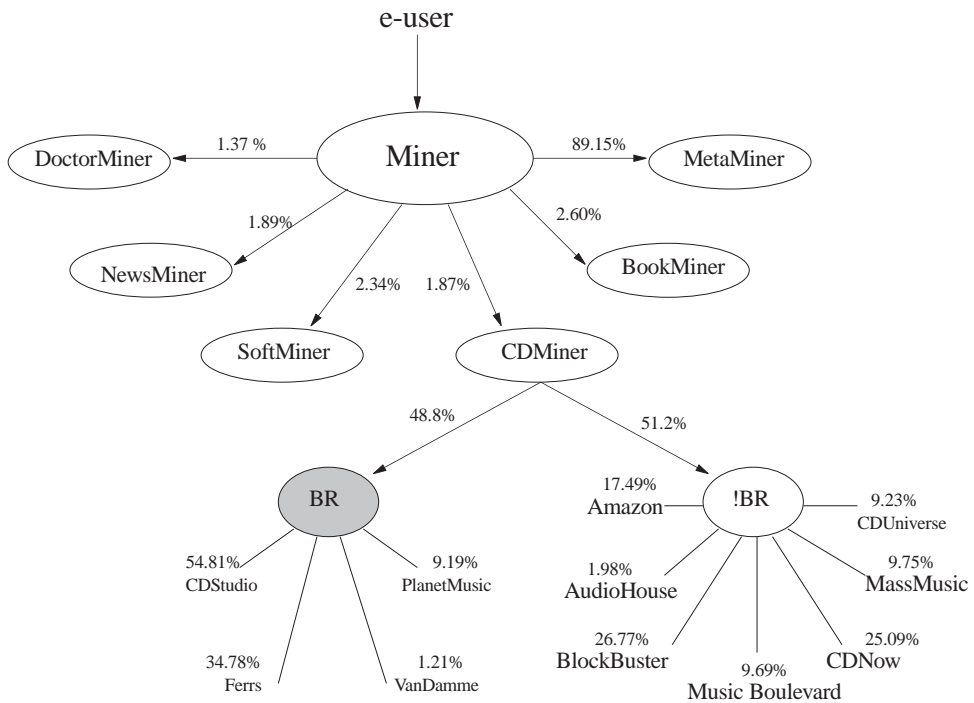


Fig. 4. CDMiner customer behavior graph.

Table 5
BookMiner performance results

Bookstore	Availability (% of requests)	Response time (s)	Book hit ratio (%)
Barnes & Noble	95.55	25.4	18.5
Bookstacks	84.75	8.1	22.0
BookPool	99.50	10.4	4.7
McGraw Hill	99.20	28.0	4.3
O'Reilly	100.00	12.7	4.6
Prentice Hall	100.00	7.1	7.2
iBS	100.00	17.1	13.9
Amazon	99.23	13.0	19.1
Booknet	98.27	12.5	49.3
Campus	100.00	2.0	7.2
Cultura	100.00	14.3	33.6
Siciliano	76.12	24.8	69.4
Sodiler	100.00	11.4	38.4
Tempo Real	100.00	12.5	11.5
Loyola	100.00	8.9	56.0
artepaubrasil	100.00	8.9	55.7
BookMiner	100.0	48.5	

Table 6
CDMiner performance results

Music store	Availability (% of requests)	Response time (s)	CD hit ratio (%)
Amazon	98.73	17.271	22.22
AudioHouse	100.00	11.259	8.31
BlockBuster	100.00	6.763	22.14
CDUniverse	97.47	34.994	23.11
CDNow	98.95	20.099	33.44
MassMusic	94.86	41.356	26.26
MusicBoulevard	97.27	18.550	16.52
Ferrs	100.00	5.124	57.55
PlanetMusic	100.00	4.850	31.89
VanDamme	100.00	1.955	5.07
CDStudio	100.00	6.824	44.73
CDMiner	100.00	41.698	

request response time is the interval of time needed for receiving a response from the server. Tables 5 and 6 show availability and elapsed response time of the registered stores that are queried by BookMiner and CDMiner, respectively. We note that almost all stores exhibit a good level of availability no matter where they are located. On the other hand, the same tables show a high variance for response times. Average elapsed response time of national stores is lower than the same time of international stores. We con-

jecture that this phenomenon is a consequence of the heavy traffic on the international links between Brazil and US. One remarkable exception is BlockBuster, that answers as fast as any Brazilian music store.

It is worthwhile to look at the influence of those metrics in the measurements we obtained for the brokers. For example, Siciliano bookstore does not exhibit a good service level indicator. It has the lowest availability (23.88% of the queries timed out as shown in Table 5) among the Brazilian bookstores. However, Siciliano is the Brazilian bookstore that attracted the largest portion of the Brazilian customer community. This apparent contradiction can be explained by the influence of the ‘brand’ on the shoppers. Siciliano is a well established company, having many bookstores in the main cities of Brazil, which somehow makes the company familiar to customers, even on the Internet.

We define another metric called ‘book hit ratio’ (BHR) that represents the number of times that a bookstore suggests at least one title in response to a customer request over the total number of requests sent to the bookstore. Looking at Table 5, it is evident that Brazilian bookstores are more effective in finding in their selection the books requested by Brazilian customers. The BHR of Brazilian bookstores is higher than the BHR of the global bookstores. This fact stems from cultural factors such as English proficiency and local interests. Around 50% of Brazilian Internet users do not know English. Also, Brazilian bookstores have much larger selection of books on topics that are part of the Brazilian culture [1] than global bookstores.

Regarding music stores, we define a similar metric called ‘CD Hit Ratio’ (CHR) that represents the percentage of requests that a music store suggests at least one CD in response to a customer request. Looking at Table 6, we note that all Brazilian music stores but VanDamme (which is specialized on *new age*) presented a higher CHR than international music stores, which is explained by requests to Brazilian artists. However, notice that the CHRs are smaller than BHRs for national stores, confirming the smaller influence of language issues on music preferences.

A final observation regards the percentage of click-throughs that turned into sales. We analyzed sale reports from three stores (i.e., amazon, CDNow and Booknet) and found that 8% of the click-

throughs became sales of books and 3% of the click-throughs turned into CD sales. This behavior has been observed before [10], that is, consumers are less likely to buy CDs on electronic stores than they are to buy books. Recently, [14] presented similar results, stating that only 5% of the visits to e-commerce sites are to buy. These results show that brokers are more efficient than some advertisement mechanisms, such as banners (whose estimated click-through ratio is only 1%).

4. Case study: efficiency of a non-English e-broker

The efficiency of the results provided by a broker could be assessed by the percentage of customers that is driven to each of the registered stores. In Section 3, we saw that BookMiner turned 8% of the click-throughs into book sales. Now, we want to answer the following question: What are the reasons that motivated customers to shop on the stores pointed out by the brokers? Based on a first intuition, we would say that the percentage of click-throughs for a given store is proportional to the book hit ratio and availability, and inversely proportional to the response time. However, looking at the data obtained from the logs, we found different observations. We did a regression analysis on the data presented in Section 3 and found that average response time is not correlated to click-through. Then, we did other correlation tests and found that the number of click-throughs for each store is strongly influenced by factors such as book hit ratio, price, brand and regional characteristics, represented by language, currency, logistics and customs. We examined the logs from the operation of BookMiner for two days and assessed what factors were influential on the customer preference.

4.1. Bookstore selection on click-through percentages

The availability of a large variety of products is a key issue in the relationship between customers and companies on the Internet. Bookstore selection is directly related to the book hit ratio metrics, that represents the number of times that a bookstore suggests

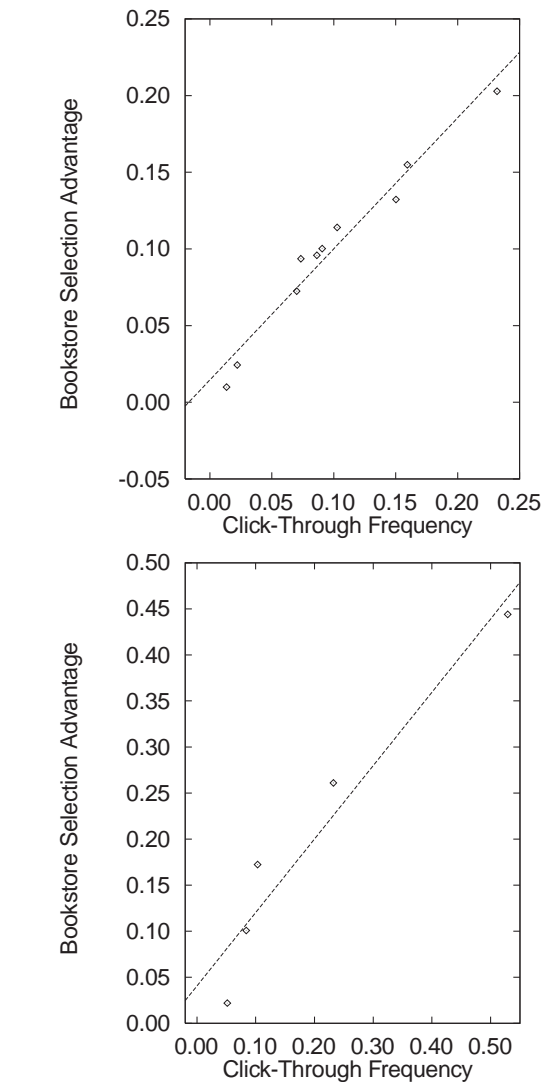


Fig. 5. Bookstore selection influence: national and international.

at least one title in response to a customer request over the total number of requests sent to the store. Fig. 5 shows that there exists a strong relation between the availability of titles and the click-through percentages. The larger the selection of a given store, the greater the click-through percentage for the store.

Bookstore selection quantifies the diversity of titles offered by a bookstore. It is calculated in a per-click basis: for every bookstore that offers the desired title, we add the inverse of the number of

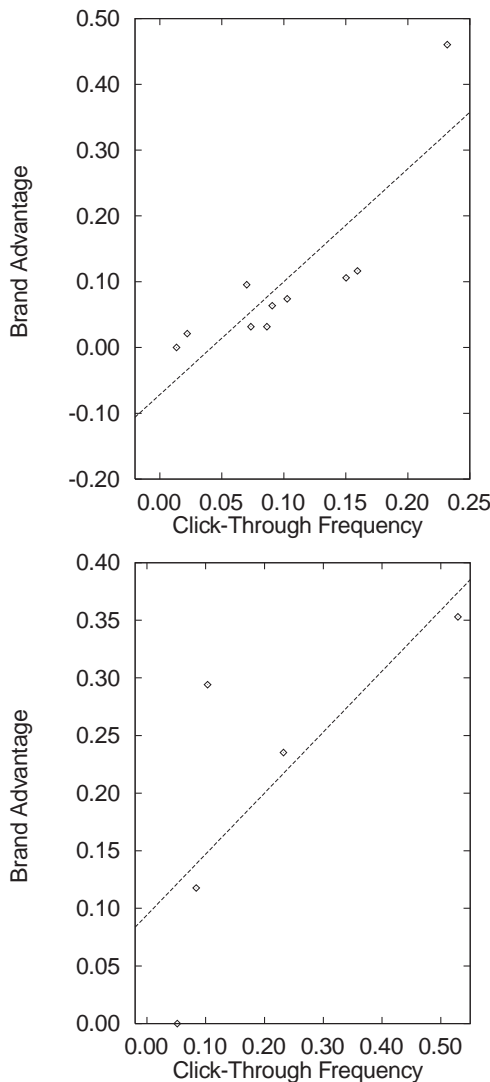


Fig. 6. Brand influence: national and international.

offering bookstores to its bookstore selection. The percentages shown in the graphs of Fig. 5 are the relative weight of each bookstore considering the overall bookstore selection observed.

4.2. Brand influence on click-through percentages

Trust is a fundamental issue in the relationship between customers and online stores. Trust in the electronic market is many times associated with the traditional concept of retail brand, that identifies

the e-commerce company that is responsible for the customer relationship in a electronic transaction. In this case study, we viewed the factor 'brand' as the percentage of those requests where the bookstore offer was clicked despite its price was not the minimum among the several offers or within 10% of the minimum price. In this case, we conjectured that the bookstore choice was driven by the 'bookstore brand'. Fig. 6 shows that there exists in our data a correlation between brand and click-through percentage for both national and international bookstores. In the previous section we observed the importance of the brand factor when analyzing the number of click-throughs to the Siciliano bookstore.

4.3. Price advantage on click-through percentages

Price advantage is calculated as the percentage of requests associated with each bookstore that offered the best price, among all requests chosen by price. Fig. 7 displays the graphs of price advantage versus click-through percentages for both national and international bookstores. In the international market (see Fig. 7), we note strong evidences towards the relevance of the price advantage as a factor that drives consumers. We can observe the influence of price by comparing three bookstores that are specialized in technical books: Bookpool, O'Reilly, and McGraw-Hill. The book hit ratio of them is around 4.5% but their click-through percentages vary significantly (5.4%, 1.43%, and 3.25%, respectively). Bookpool attracts more consumers because its prices are usually lower than other bookstores.

5. Concluding remarks

This paper shows an overview of the Miner Family architecture and implementation. We also present a quantitative study of the behavior of two large non-English e-brokers. We characterize the workload of the Miner Family and focus on the behavior of the BookMiner and the CDMiner, two of its brokerage services. Using these quantitative results, we proposed and analyzed an efficiency model for e-brokers.

Based on the statistics shown in the paper we found interesting observations about e-commerce.

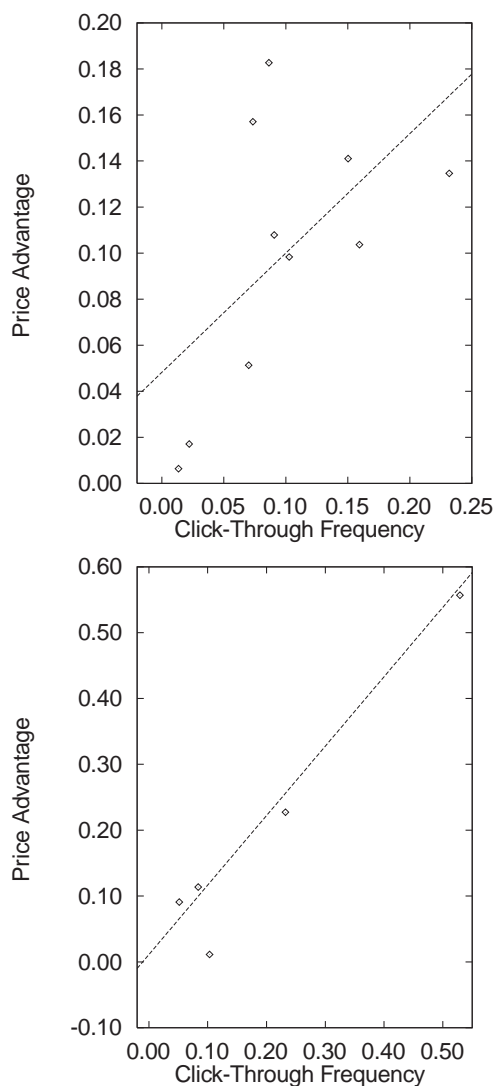


Fig. 7. Price influence: national and international.

First, we note that 76% of the Brazilian customers prefer Brazilian bookstores. Network infrastructure also affects the customer behavior. Average elapsed response time of national bookstores is lower than international bookstores. We conjecture that this phenomenon is a consequence of the heavy traffic on the international links between Brazil and US. Language and social aspects play a major role in the behavior of e-commerce customers. Brazilian bookstores are more effective to attract customers, not because they offer more titles, but because they offer Portuguese

books, which are far more popular. Also, customs, currency conversion and delivery logistics help local bookstores. The behavior of CD consumers is quite different from book consumers. The average response time for Brazilian music stores is also lower than global music stores, but customers visit the music stores regardless of their location. We also proposed a model that quantifies the efficiency of the results provided by brokers in the electronic marketplace as a function of bookstore selection, price and brand.

Although the Web opens a company to a global market, our findings clearly indicate that e-commerce is strongly tied to regional issues, such as language, national customs and regulations, currency conversion and logistics. Also, the Internet infrastructure, mainly the intercontinental links, hinders a consistent performance and affects the user behavior. Our future work will focus on enhancing the quantitative analysis of the behavior of e-commerce users by extending the presented model and coming up with other models that describe workloads of e-commerce components, such as portals, brokers and merchants. Moreover, we would like to answer questions such as what regional features should be present in a portal site considering cultural and language characteristics.

References

- [1] V. Almeida, M. Cesário, R. Fonseca, W. Meira Jr. and C. Murta, The influence of geographical and cultural issues on the cache proxy server workload, *Computer Networks and ISDN Systems* 30 (1998) 601–603.
- [2] Y. Bakos, Reducing buyer searching costs: implications for electronic marketplaces, *Management Science* 43(12) (1997).
- [3] F. Cheong, *Internet Agents Spiders, Wanderers, Brokers and Bots*, New Riders, 1996.
- [4] D. Dreilinger, Savvysearch, main page, <http://guaraldi.cs.colostate.edu:2000/>
- [5] Express, main page, <http://www.express.infoseek.com>
- [6] Inktomi, main page, <http://www.inktomi.berkeley.edu>
- [7] Jango, main page, <http://jango.excite.com>
- [8] Jungle, main page, <http://www.jungle.com>
- [9] M. Koster, Guidelines for robot writers, <http://web.nexor.co.uk/mak/doc/robots/guidelines.html>, 1996.
- [10] M. Krantz, The Internet economy, *Time Magazine*, July 20 (1998) 28–35.
- [11] S. Lawrence and C. Giles, Searching the World-Wide Web, *Science* 280 (5360) (1998) 98.

- [12] D. Menascé and V. Almeida, *Capacity Planning for Web Performance — Metrics, Models, & Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [13] Miner family, main page, <http://miner.com.br>
- [14] J. Nielsen, Why people shop on the Web, <http://www.useit.com/alertbox/990207.html>, February 7th, 1999.
- [15] E. Selberg and O. Etzioni, Multi-service search and comparison using the metacrawler, in: Proc. 4th Int. World-Wide Web Conference, 1995.
- [16] A. Segev, D. Wan and C. Beam, Designing electronic catalogs for business value: Results from the commercenet pilot, Technical Report Working Paper CITM-WP-1005, Fischer Center for Information Technology — U.C. Berkeley, October 1995.
- [17] Universo online, main page, <http://www.uol.com.br>
- [18] Yahoo!, main page, <http://www.yahoo.com>
- [19] Yahoo shopping guide, main page, <http://shopguide.yahoo.com>



Virgílio Almeida is a Professor of Computer Science at the Federal University of Minas Gerais, Brazil. He holds a PhD in computer science from Vanderbilt University and held visiting faculty and research positions at Boston University and Xerox PARC. He has published extensively in the area of distributed system and World Wide Web performance and is the author of the book *Capacity Planning for Web Performance: Metrics, Models, and Methods*, published by Prentice-Hall, 1998.

Capacity Planning for Web Performance: Metrics, Models, and Methods, published by Prentice-Hall, 1998.



Wagner Meira Jr. is a research associate in the Department of Computer Science at the Federal University of Minas Gerais, Brazil. He holds a PhD in computer science from the University of Rochester. His current interests are on performance analysis and modeling of parallel and distributed systems, in particular the World Wide Web.



Victor Fernando Ribeiro received his Bachelor degree in Applied and Computational Mathematics in 1992, at the Universidade Estadual de Campinas. He started his career in 1992 at the Wire Drawing Division of Companhia Siderúrgica Belgo-Mineira, developing several studies in the fields of Operations Research and Industrial Engineering. He was technology analyst and assistant to the Belgo Mineira

Company's CIO, during 1996 and 1997. In 1998 he received his MSc degree in Computer Science at the Federal University of Minas Gerais, where he designed the Miner Family. Presently, he is co-founder and the Chief Operating Officer of Miner Technology Group.



Nivio Ziviani is a Professor of Computer Science at the Federal University of Minas Gerais (UFMG) in Brazil, where he heads the Laboratory for Treating Information (LATIN). He received a PhD in Computer Science from the University of Waterloo, Canada, in 1982. He currently coordinates a four years project on Web and wireless information systems (called SIAM) financed by the Brazilian Ministry of

Science and Technology. He is co-founder of the Miner Technology Group, owner of the Miner Family of agents to search the Web. He is the author of several papers in journals and conference proceedings covering topics in the areas of algorithms and data structures, information retrieval, text indexing, text searching, text compression, and related areas.